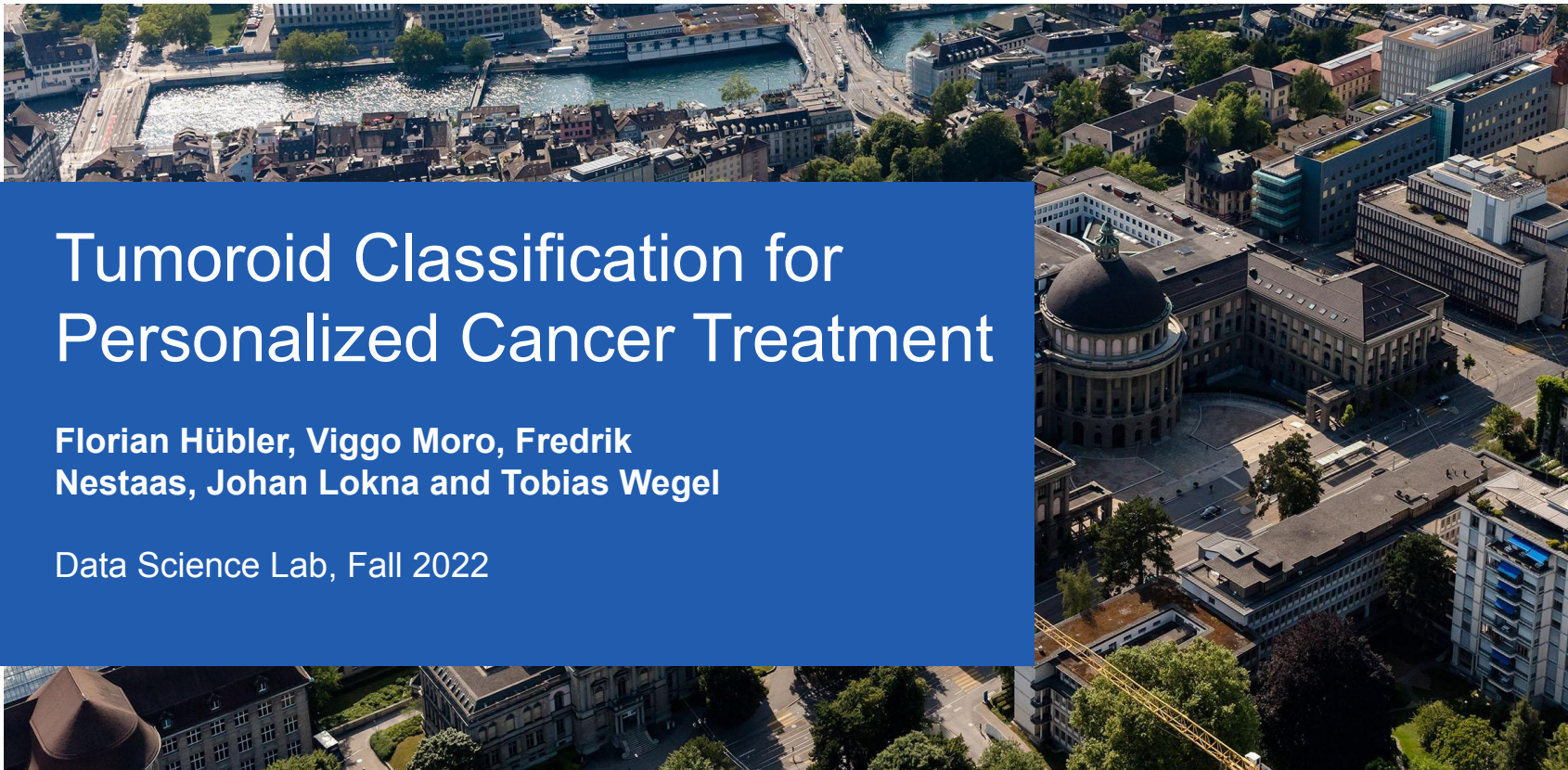


Tumoroid Classification for Personalized Cancer Treatment

Florian Hübler, Viggo Moro, Fredrik
Nestaas, Johan Lokna and Tobias Wegel

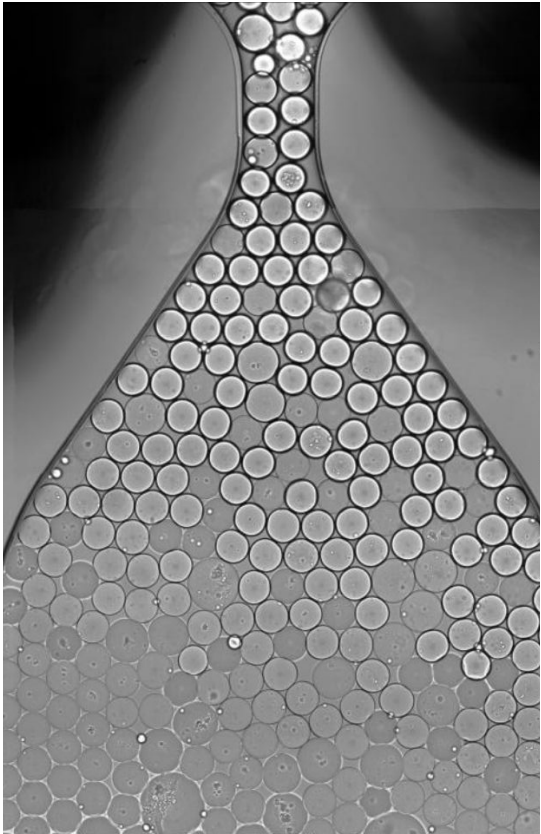
Data Science Lab, Fall 2022



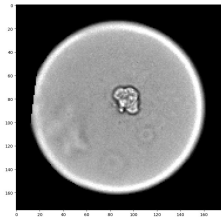
Our role as data scientists

- Find a way to automate the classification of tumor droplets.
- Use deep learning to achieve this.
- Make use of manually annotated tumor droplets for training.
 - Training data \implies model
 - testing data \implies confusion matrix
 - prediction data \implies drug response
- Draw conclusions about response of different drugs.
- Jupyter notebook as interface to use our code

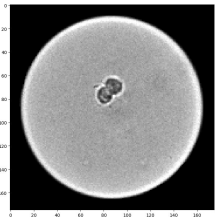
Data Preprocessing



1

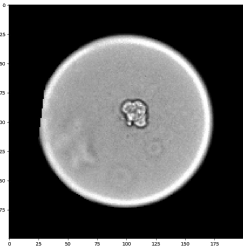


,

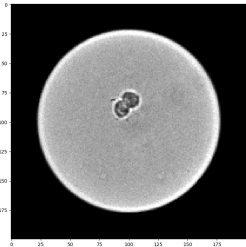


, ...

2

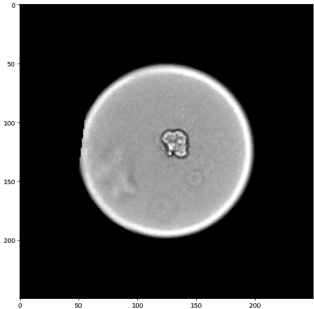


,

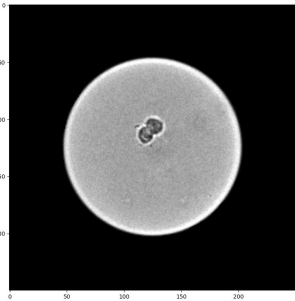


, ...

3



,

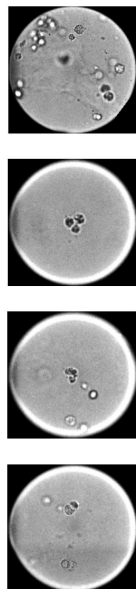


, ...

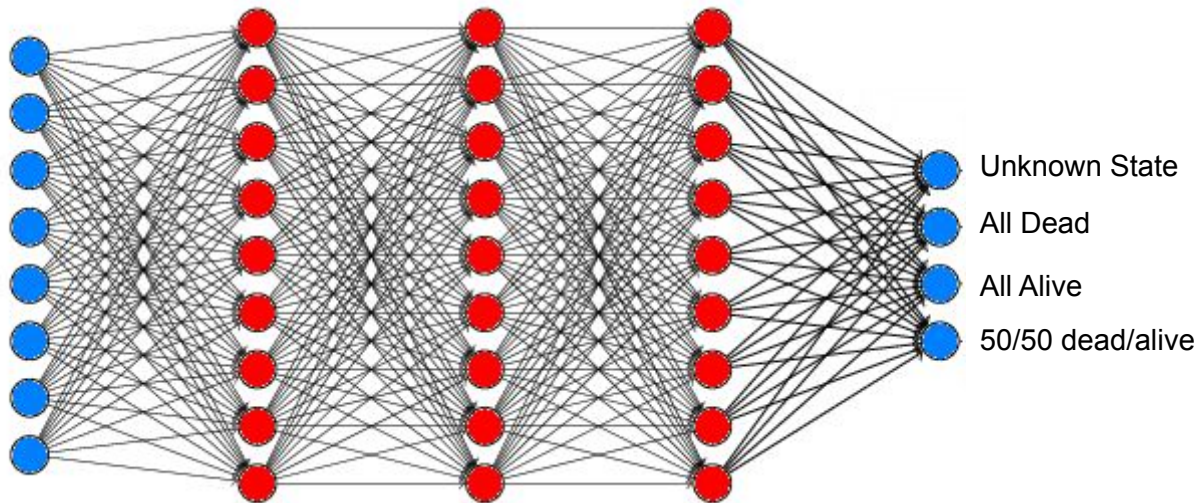
Always use the same
preprocessing for training
and prediction!

Demonstration of preprocessing in the Notebook

Model & Training



Droplets

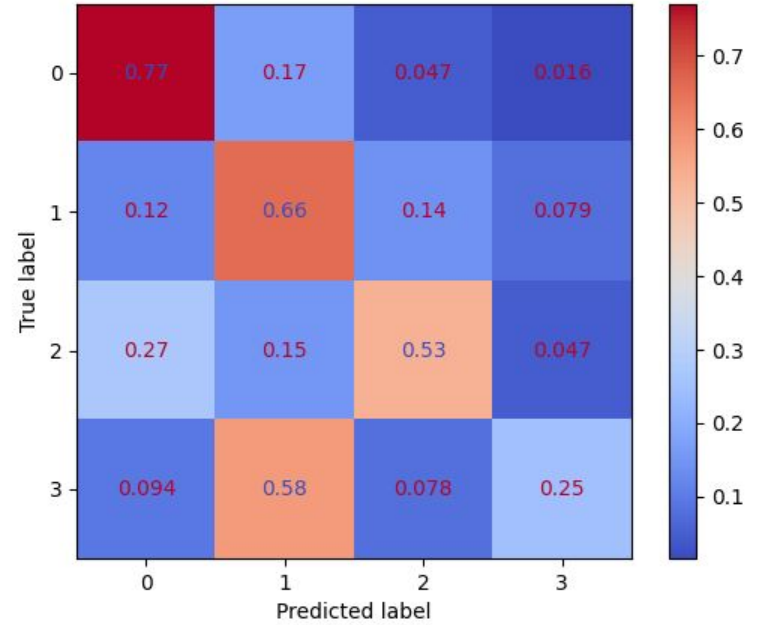
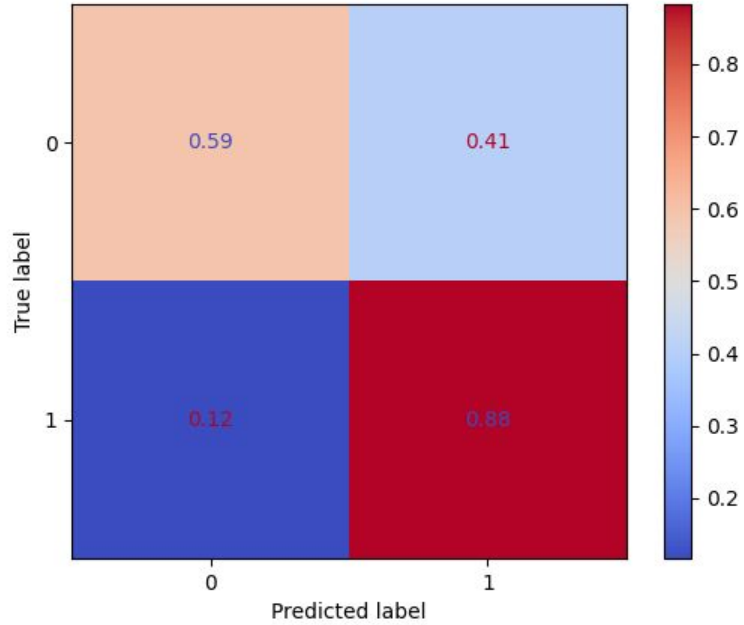


ResNet34

- Unknown State
- All Dead
- All Alive
- 50/50 dead/alive

Weighted Cross-Entropy

Results



Demonstration of training in the Notebook

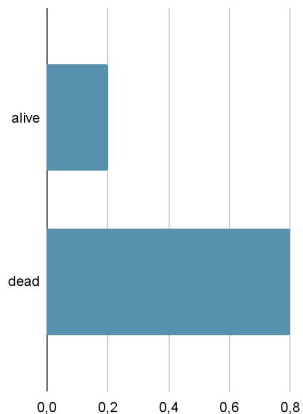
Key insight:

We only care about *overall* drug effect per patient

Estimating Label Distribution per Drug

With the right statistical model:

label distribution of drug

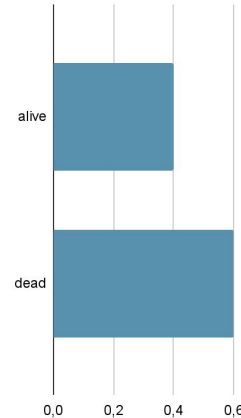


=

confusion matrix

estimate on test set

prediction distribution of drug
from model

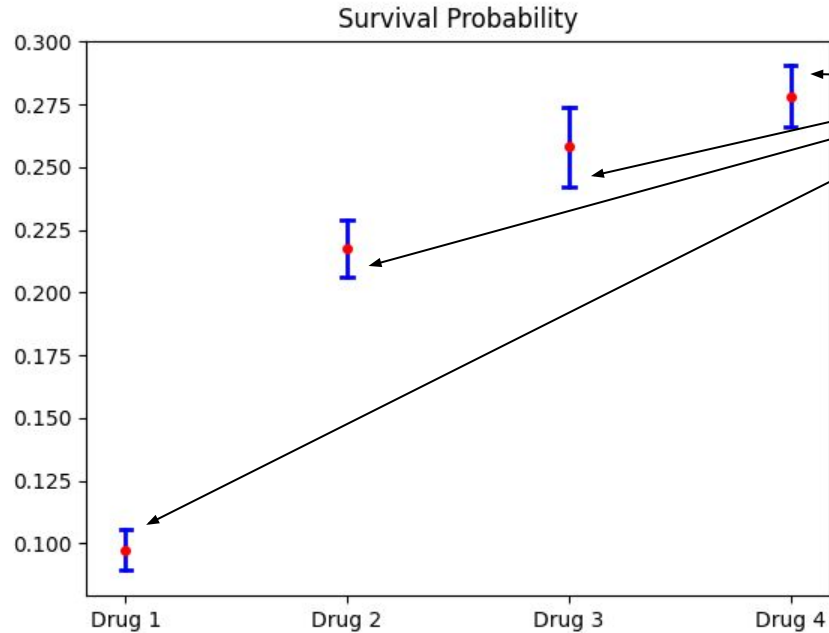


*

estimate during inference

How good are the estimates?

Different Drugs, Different Survival



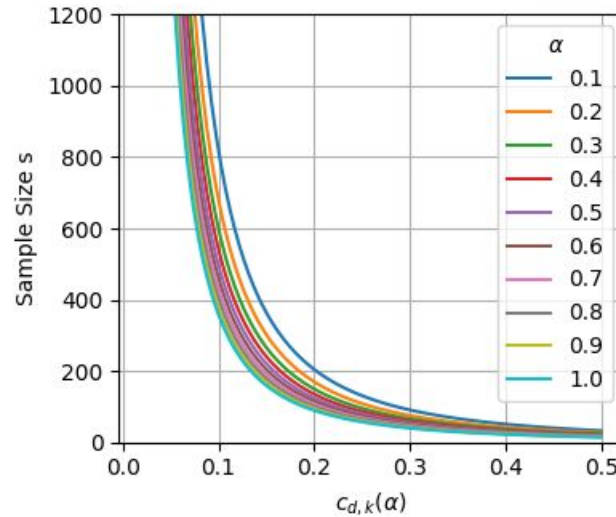
$(1-\alpha)$ - confidence intervals for class alive:

1. Derived from concentration inequalities.
2. Hold uniform over the drugs.
3. If they do not overlap:
With probability $(1-\alpha)$ we have found the best drug!

Note: this plot is a sketch and was not computed using actual data.

More samples means higher confidence

Deviation vs. sample size for different confidences for $C = 2$



A word of caution

- These confidence bounds depend on mathematical assumptions.
- If those assumptions do not hold, you cannot trust the results.
- Main advantage of this approach:
We can get some idea of the uncertainty in our estimates.

Demonstration of prediction in the Notebook

Thank you!
Any questions?